LINEAR MODELS OF MULTIDIMENSIONAL FIELDS IN THE KARHUNEN-LOEVE BASIS

K.T. Protasov

Institute of Atmospheric Optics,
Siberian Branch of the Academy of Sciences of the USSR, Tomsk
Received November 9, 1988

An iteration algorithm for the approximate solution of the problem of determining the Karhunen-Loeve basis using experimental data within the framework of a linear model of random vector fields of the vector argument is suggested. An example of such fields can be seen in an ensemble of multizonal images.

Linear models for representing data in orthogonal bases are widely used in the processing of multidimensional experimental data. The most preferable among all these is the Karhunen-Loeve basis (known in the literature on meteorology, hydrology, oceanology, atmospheric and ocean physics as the basis of empirical orthogonal functions);[1–5] in this case the approximating series has the least number of components, while retaining a high accuracy of approximation of the data.[6,8]

Tendencies which have appeared of an increasing dimensionality of the recorded data (in particular, the appearance of a multizonal aerospace survey) and the problems of Joint data processing arising from it add urgency to the problem of creating linear models of multidimensional observations described by means of random vector fields. In this connection, this paper considers the quite general problem of representing vector fields of a vector argument, as well as the problem of finding a corresponding Karhunen-Loeve basis based on the experimental data, and an iterative algorithm is suggested for the approximate solution of the above problem. We assume that a random (for simplicity and without loss of generality, centered) vector field $\vec{\xi}(\vec{u}) = (\xi^1(\vec{u}), ..., \xi^s(\vec{u}))^T$ of a vector argument $\vec{u} = (u^1, ..., u^\nu)^T$ (where s and ν are the dimension of the function $\vec{\xi}(\cdot)$ and of the argument $\vec{u}$, respectively, and T is the transposition symbol) is represented in its domain of definition

$$D = \{\mathbf{u} : u_a^i \leqslant u^i \leqslant u_b^i, \ i = 1, ..., \ \nu\}$$

by a set of N realizations $\vec{\xi}_1(\vec{u}), ..., \vec{\xi}_n(\vec{u})$.

Let us represent a vector field In the following (generally speaking, non-unique)[7] form:

$$\vec{\xi}(\vec{u}) = \lim_{\kappa \to \infty} \sum_{i=1}^{\kappa} X^i \vec{\Phi}_i(\vec{u}), \qquad (1)$$

where the limit Is understood in the sense of non convergence in the space of realizations of a random vector of the vector field; $\{\vec{\Phi}_1(\vec{u})\}_k$ are the basis vector functions of the vector argument. Random coefficients $\{X^1\}_k$ are determined by the minimum condition of the mean square weighted deviation

$$E_k^2 = M \left| \psi^{\frac{1}{2}}(\vec{u}) \cdot \left(\vec{\xi}(\vec{u}) - \sum_{i=1}^{k} x^i \vec{\phi}_i(\vec{u}) \right) \right|^2. \qquad (2)$$

where M is the mathematical expectation operator, $\Psi(\vec{u})$ is a scalar weight function, non-random on D; and $\| \cdot \|$ is the Euclidean norm in the observation space.

If one imposes upon the basis functions $\{\vec{\Phi}_1(\vec{u})\}_k$ the orthonormality conditions

$$\left[\vec{\phi}_i, \ \vec{\phi}_j \right]_\psi = \int_D \psi(\vec{u}) \ \vec{\phi}_i^T(\vec{u}) \ \vec{\phi}_j(\vec{u}) \ d\vec{u} = \delta_{ij}. \qquad (3)$$

where δ_{ij} ($i, j = 1, ..., k$) is the Kronecker symbol, $d\vec{u} = du^1 \times ... \times du^\nu$, and $(\cdot, \cdot)_\psi$ is the symbol of the weighted scalar product, then the representation coefficients $\{X^1\}_k$ that minimize expression (2) have the form

$$x^i = \left[\vec{\xi}, \ \vec{\phi}_i \right]_\psi = \int_D \psi(\vec{u}) \ \vec{\xi}^T(\vec{u}) \ \vec{\phi}_i(\vec{u}) \ d\vec{u}, \ i=1, ..., k. \qquad (4)$$

The existence of a limit of the series (1) and of a complete orthogonal sequence of basis functions $\{\vec{\Phi}_1(\vec{u})\}_\infty$ is assured by considering only those processes $\vec{\xi}(\vec{u})$ that satisfy the following condition: for any fixed set of the values of the components of a vector $\vec{u} \in D$, $M[\vec{\xi}^T(\vec{u})\vec{\xi}(\vec{u})] < \infty$. One can also find basis functions $\{\vec{\Phi}_1(\vec{u})\}_k$ from the minimum condition of the mean square test of the goodness criterion (2) of the approximation of the vector field $\vec{\xi}(\vec{u})$ by the series (1) truncated to k terms.

A solution of the above variational problem for a conditional (in the sense of constraints (3) imposed upon the functional (2) via the Lagrange multipliers) extremum leads to a homogeneous Fredholm integral equation of the second kind

$$\int_D \psi(\vec{v})\, M\left[\,\vec{\xi}(\vec{u})\,\vec{\xi}^T(\vec{v})\,\right]\,\vec{\phi}(\vec{v})\,d\vec{v} = \lambda\,\vec{\phi}(\vec{u}). \tag{5}$$

where λ is a Lagrange multiplier, and all the basis function indices are omitted because of the equivalence of all the equations.

The desired basis $\{\vec{\Phi}_1(\vec{u})\}_k$, corresponding to the k greatest eigenvalues $\{\lambda_1\}_k$, is found by solving Eq. (5); however, in the general case this problem is not easy.

Having at our disposal a set of N realizations $\vec{\xi}_1(\vec{u}), ..., \vec{\xi}_N(\vec{u})$ (a sample of size N), which characterizes the general set of all realizations generated by a random field $\vec{\xi}(\vec{u})$ "sufficiently completely," it is natural to make use of the following sample estimate of the correlation function

$$M\left[\,\vec{\xi}(\vec{u})\,\vec{\xi}^T(\vec{v})\,\right] \cong \frac{1}{N}\sum_{j=1}^{N}\vec{\xi}_j(\vec{u})\,\vec{\xi}_j^T(\vec{v}). \tag{6}$$

In this case problem (5) simplifies substantially (the case of a degenerate kernel in Eq. (5)).[9] Indeed, by substituting Eq. (6) in the integral equation (5), we get

$$\int_D \psi(\vec{v})\sum_{j=1}^{N}\vec{\xi}_j(\vec{u})\,\vec{\xi}_j^T(\vec{v})\,\hat{\vec{\phi}}(\vec{v})\,d\vec{v} = \Delta\,\hat{\vec{\phi}}(\vec{u}). \tag{7}$$

where $\Delta = N\hat{\lambda}$, and $\hat{\lambda}$ and $\hat{\vec{\phi}}(\vec{u})$ are estimates of the corresponding quantities λ and $\vec{\Phi}(\vec{u})$.

Let us introduce the following notation:

$$\int_D \psi(\vec{v})\,\vec{\xi}_j^T(\vec{v})\,\hat{\vec{\phi}}(\vec{v})\,d\vec{v} = c^j, \quad j = 1, ..., N$$

then from Eq. (7) we obtain an expression for the basis functions

$$\hat{\vec{\phi}}(\vec{u}) = \frac{1}{\Delta}\sum_{j=1}^{N}c^j\,\vec{\xi}_j(\vec{u}). \tag{8}$$

where the coefficients $\{C^J\}_N$ are still undefined. Substituting the parameterized expression (8) of the basis function $\hat{\vec{\phi}}(u)$ in Eq. (7), we obtain the equality

$$\frac{1}{\Delta}\sum_{j=1}^{N}\vec{\xi}_j(\vec{u})\sum_{l=1}^{N}c^l\int_D \psi(\vec{v})\,\vec{\xi}_j^T(\vec{v})\,\vec{\xi}_l(\vec{v})\,d\vec{v} = \sum_{j=1}^{N}c^j\,\vec{\xi}_j(\vec{u}). \tag{9}$$

Let us calculate the scalar product in this expression $(\vec{\xi}_j, \vec{\xi}_l)_\psi$ by means of the random field realizations

$\vec{\xi}_1(\vec{u}), ..., \vec{\xi}_N(\vec{u})$, and denote it by a_{ij}; in what follows equality (9) takes the form

$$\sum_{j=1}^{N}\vec{\xi}_j(\vec{u})\left\{\frac{1}{\Delta}\sum_{l=1}^{N}c^l a_{jl} - c^j\right\} = 0. \tag{10}$$

By force of the linear independence of the random field realizations in the random sense and due to a property of linear independent elements in a space with a scalar product, Eq. (10) holds under the condition

$$\frac{1}{\Delta}\sum_{l=1}^{N}c^l a_{jl} - c^j = 0, \quad j = 1, ..., N. \tag{11}$$

Using matrix notation, Eq. (11) can be rewritten in the form

$$(a_{ij})\,c = c\vec{\Delta}. \tag{12}$$

where $C = (C^1, ..., C^N)^T$, (a_{ji}) is the $N{\times}N$ Gram matrix, and $\vec{\Delta} = (\Delta^1 \delta_{ij})$.

Thus, after defining the structure of the basis functions in the form of a linear combination of realizations of the random process, the coefficients of these linear combinations fall out as a result of solving the full eigenvalue problem for a positive definite Gram matrix (a_{ji}) of order N, and this problem is quite realizable in practice by means of algebraic numerical methods.

It is easy to show by substituting expression (8) in Eq. (3) that to normalize the functions $\left\{\hat{\vec{\Phi}}_1(u)\right\}_k$, taking into account the obtained values $\{C^l\}_N$, it is necessary to change Δ to $\Delta^{1/2}$ in Eq. (9).

The construction of linear models (1) with the help of the optimal (in the mean-squared sense) Karhunen-Loeve basis requires solving equations (12), as a rule, by numerical methods. However, the difficulties of its practical realization limit the extensive application of that basis because in that case it is necessary to solve the full eigenvalue problem for positive definite matrices whose order exceeds 10^2. This forces one to relinquish direct methods of solving the problem of finding a Karhunen-Loeve basis (Eq. (12)) and to construct iterative algorithms which, while decreasing the number of operations needed to obtain an approximate result, lead only asymptotically to the optimal solution. Another advantage of iterative algorithms is the fact that they allow "the most important" basis functions, the number of which may be not large, to be found in the first place.

One of the ways to surmount the above difficulty at the price of rejecting the optimality in the mean-square sense is represented by the algorithm of constructing an adapted basis,[10] using the idea of orthogonalizing a sequence of linearly Independent functions under the condition that the choice of the next function is subject to a certain criterion. In this case a uniform approximation of the process is performed by

a linear manifold of a low dimension. The iterativity of this basis construction procedure allows one to use an algorithm for transforming spaces of large dimensions up to 10^5.

In what follows an iterative algorithm is proposed for constructing a basis adapted in the mean-square sense (AMS basis). In the given case the choice of the next basis function is based on the minimization of a mean-square goodness criterion (2).

Let us approximate the realizations of the initial description of $\vec{\xi}(\bar{u})$ by means of elements of the linear span G_k, given by the orthonormal basis $\{\bar{\varphi}_j(\bar{u})\}_k$, as follows: take for the next basis function $\bar{\varphi}_j(\bar{u})$ ($j = 1$, ..., k) that one amongst s_j orthonormal functions $\vec{\varphi}_{s_j}(\bar{u})$, obtained via the Gram-Schmidt orthogonalization procedure[12] applied to the sample functions $\vec{\xi}_1(\bar{u})$, ..., $\vec{\xi}_M(\bar{u})$

$$\vec{\phi}_{s_j}(\bar{u}) = \frac{\vec{\phi}'_{s_j}(\bar{u})}{|\vec{\phi}'_{s_j}(\bar{u})|},$$

$$\vec{\phi}'_{s_j}(\bar{u}) = \vec{\xi}_{s_j}(\bar{u}) - \sum_{i=1}^{j-1} \left(\vec{\xi}_{s_j} \cdot \vec{\phi}_i \right) \vec{\phi}_i(\bar{u}). \tag{13}$$

for which

$$\lambda_j = M\left[(\vec{\xi}, \vec{\phi}_j)^2 \right] = \max_{(s_j)} M\left[(\vec{\xi}, \vec{\phi}_{s_j})^2 \right].$$

$$s_j = 1, \ldots, N; \quad j = 1, \ldots, k; \quad k \leq N \tag{14}$$

where as the mathematical expectation we use a sample estimate associated with Eq. (6).

The process of finding basis functions is completed at the k-th step as soon as one reaches a given accuracy E_k^2 in the approximation of the random vector field by a linear combination of k basic elements from G_k. Then, from the theorem on projecting the elements $\vec{\xi}(\bar{u})$ of a Hilbert space onto G_k, we have

$$\vec{\xi}(\bar{u}) = \sum_{j=1}^{k} x^j \vec{\phi}_j(\vec{u}). \tag{15}$$

where $\{X^j\}_k$ is a set of random numbers defined by the formula

$$X^j = (\vec{\xi}, \vec{\phi}_j), \quad j = 1, \ldots, k. \tag{16}$$

The accuracy of the approximation in expression (15) is determined (in practice, estimates of $M[\cdot]$ over the same sample are used) as follows:

$$E_k^2 = M\left[(\vec{\xi}, \vec{\xi}) \right] - \sum_{j=1}^{k} \lambda_j. \tag{17}$$

where λ_j are in decreasing order ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$) by virtue of their construction, and, to the maximum extent allowed by the set $\{\vec{\xi}_j(\bar{u})\}_N$, they "exhaust" the rms error E_k^2 of the approximation to the ensemble $\{\vec{\xi}(\bar{u})\}$ by a linear manifold from G_k. As is well known, an optimal (in the rms sense) basis $\{\vec{\Phi}_j(\bar{u})\}_k$ with the corresponding eigenvalue spectrum $\{\lambda_j\}_k$ of the Karhunen-Loeve expansion is found by optimizating criterion (17) over $\{\bar{\varphi}_j(\bar{u})\}_k$, taking into account the orthonormalizability of the latter functions. The problem of successive maximization of the positive definite quadratic form (14) on unit spheres within subspaces orthogonal to the functions $\{\bar{\varphi}_j(\bar{u})\}_k$, obtained via procedure (13), $j = 1, \ldots, k$, $\bar{\Phi}_0 \equiv 0$, leads to the same basis $\{\Phi_j(u)\}_k$. The latter circumstance indicates that the AMS basis $\{\varphi_j(\bar{u})\}_k$ obtained using algorithm (13), (14) becomes a Karhunen-Loeve basis asymptotically with increasing N under certain assumptions on the ensemble $\{\vec{\xi}(\bar{u})\}$. Indeed, the algorithm which we have in fact considered is based on stochastic principles of an extremum search[11] with the only peculiarity being that the "test" functions in this case are elements of a sample. Therefore it is necessary to ensure convergence of the 'search procedure that the sample value distribution function be positive in the "direction" of searched-for solutions. If this fact is not established a priori, then one can Judge the quality of the obtained solution by the magnitude of the estimate of the error of approximation (17):

$$\bar{E}_k^2 = \frac{1}{N} \sum_{j=1}^{N} \left[(\vec{\xi}_j, \vec{\xi}_j) - \sum_{i=1}^{k} (\vec{\xi}_j, \vec{\phi}_i)^2 \right].$$

It should be noted that algorithms obtained for continuous fields remain valid for fields given by individual readings in a discrete regular or a stochastic (but fixed) observation network; however, in these cases the integration is replaced to the corresponding summation over a set of points at which the field realizations are recorded. It is not complicated to perform all of the modifications of the algorithms connected with representation of the fields in a domain D with variable boundaries, moving boundaries, or to distinguish in a linear model (1) a temporal variable of the form

$$\vec{\xi}(\vec{u}, t) = \sum_{i=1}^{k} x^i(t) \vec{\phi}_i(\vec{u}).$$

REFERENCES

1. N.A. Bagrov, Meteorol. Gidrol., No. 12, 5 (1978).
2. M.I. Fortus, Meteorol. Gidrol.,No. 4, 113 (1980).
3. A.O. Artem'ev, Okeanologiya **27**, No. 2, 204 (1987).

4. M.I. Fortus, Okeanologiya **11**, No. 11, 1107 (1975).

5. A.M. Obukhov. Izv. Akad Nauk SSSR, Ser. Geofiz., No. 3, 432 (1960).

6. S. Watanabe, in: *Transactions of the Fourth Prague Conference on Information Theory, Statistical Decision Functions. and Random Processes.* Prague, September. 1965 (Publ. House of the Czechoslovak Acad, of Scl., Prague. 1967).

7. H.L. Van Trees. *Estimation and Modulation Theory. Part 1. Detection, Estimation, and Linear Modulation Theory* (Wiley. New York. 1968).

8. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press, New York, 1972).

9. A.F. Verlan' and V.S. Sizikov, *Computer Methods of Solving Integral Equations* (Naukova Dumka, Kiev, 1978).

10. Yu.I. Neimark, ed., *Pattern Recognition and Medical Diagnostics* (Nauka, Moscow, 1972).

11. I.M. Sobol', *Numerical Monte Carlo Methods* (Nauka, Moscow, 1973).

12. G.E. Shilov, *Mathematical Analysis. Special Course* (Nauka, Moscow, 1960).